

(19) 世界知的所有権機関  
国際事務局(43) 国際公開日  
2003年7月24日 (24.07.2003)

PCT

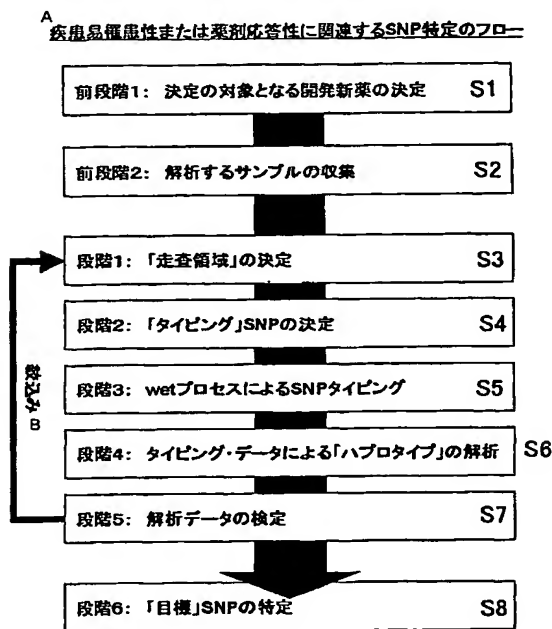
(10) 国際公開番号  
WO 03/060162 A1

- (51) 国際特許分類<sup>7</sup>: C12Q 1/68, C12N 15/09, G01N 33/50 (71) 出願人 および  
(72) 発明者: 鈴木 俊明 (SUZUKI,Toshiaki) [JP/JP]; 〒251-0015 神奈川県 藤沢市川名 849番地10号クリオ藤沢参  
(21) 国際出願番号: PCT/JP03/00261 番館 506 Kanagawa (JP).  
(22) 国際出願日: 2003年1月15日 (15.01.2003) (72) 発明者; および  
(75) 発明者/出願人 (米国についてののみ): 鎌谷 直之 (KA-MATANI,Naoyuki) [JP/JP]; 〒186-0005 東京都 国立  
(25) 国際出願の言語: 日本語 市 西2-29-89 Tokyo (JP). 田中 順治 (TANAKA,Junji)  
(26) 国際公開の言語: 日本語 [JP/JP]; 〒247-0071 神奈川県 鎌倉市 玉縄5-12-9 Kana-  
gawa (JP).  
(30) 優先権データ: (74) 代理人: 堀 城之 (HORI,Shiroyuki); 〒100-6035 東京  
特願2002-42355 2002年1月15日 (15.01.2002) JP 都 千代田区 霞が関3-2-5 霞が関ビル35階 Tokyo (JP).  
特願2002-213695 2002年7月23日 (23.07.2002) JP

[続葉有]

(54) Title: METHOD OF SPECIFYING SNP

(54) 発明の名称: SNP 特定方法



(57) Abstract: (1) It is intended to provide a technique relating to a method of specifying an SNP which comprises repeating presumption of an SNP serving as a marker and detailed typing of SNPs around the same, thus gradually narrowing down the focus to the base sequence domain in which the "target" SNP is likely contained and finally specifying the "target" SNP at a high efficiency. As Fig. 1 shows, the method of specifying an SNP comprises: (1) determining a drug to be developed which is the subject of the determination; (2) collecting samples to be analyzed; (3) determining a "scanning domain (base sequence domain)"; (4) determining "typing" SNPs; (5) SNP typing by the wet process and analyzing haplotypes based on the typing data; (6) presuming a "marker" SNP (determining the analytical data); and (7) specifying the "target" SNP (target SNP). A cycle consisting of the stage (1) to (7) is repeated as a treatment cycle.

A...FLOW OF SNP SPECIFICATION RELATING TO DISEASE RISK OR DRUG-RESPONSE  
 S1...PRESTAGE 1: DETERMINATION OF DRUG TO BE DEVELOPED (SUBJECT) S1  
 S2...PRESTAGE 2: COLLECTION OF SAMPLES TO BE ANALYZED S2  
 S3...STAGE 1: DETERMINATION OF 「SCANNING DOMAIN」 S3  
 S4...STAGE 2: DETERMINATION OF 「TYPING」 SNPS S4  
 S5...STAGE 3: SNP TYPING BY WET PROCESS S5  
 S6...STAGE 4: ANALYSIS OF 「HAPLOTYPES」 BASED ON TYPING DATA S6  
 S7...STAGE 5: CALIBRATION OF ANALYTICAL DATA S7  
 S8...STAGE 6: SPECIFICATION OF 「TARGET」 SNP S8  
 B...NARROWING DOWN

[続葉有]



(81) 指定国 (国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI 特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告書

(84) 指定国 (広域): ARIPO 特許 (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ特許

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

---

(57) 要約:

①マーカーとなるSNPの推定とその近傍SNPの詳細なタイピングを繰り返すことで、「目標」SNPが存在すると思われる塩基配列領域を段階的に絞込、最終的に「目標」SNPを効率良く特定するSNP特定方法に関する技術を提供する点にある。

第1図に示すように、本実施の形態1に係るSNP特定方法は、①決定の対象となる開発新薬の決定と②解析するサンプルの収集と③「走査領域 (塩基配列領域)」の決定と④「タイピング」SNPの決定と⑤wetプロセスによるSNPタイピングとタイピング・データによるハプロタイプの解析と⑥「マーカー」SNPの推定 (解析データの決定) と⑦「目標」SNP (目標SNP) の特定との工程を有し、①～⑦の工程をを1つのサイクル (処理サイクル) として繰り返す。

## 明細書

## SNP特定方法

## 技術分野

- 5        SNP機能解析プロセスの効率化と精度の向上、及び製薬企業などでの開発新薬の治験におけるSNP機能解析手法を応用した薬剤応答性のSNP特定方法に属する。

## 背景技術

- 10        従来の疾患易罹患性や薬剤応答性に関連するSNP (Single Nucleotide Polymorphism: 一塩基多型または一塩基多型となる座位) を特定するSNP機能解析プロセスでは、コスト上の理由によって、解析したいSNPを数十から数千程度の数(個所)に絞込んでからwetプロセス(注1)であるSNPのタイピングを行っている。
- 15        第7図は、従来のSNP機能解析のプロセス・フローを示す図である。第7図に示すように従来のSNP機能解析においては、前段階A(決定の対象となる開発新薬の決定)、前段階B(解析するサンプルの収集)、段階A(タイピングSNPの決定)、段階B(wetプロセスによるSNPタイピング)、段階C(データの解析)、段階D: (「目標」SNPの特定) が順次行われる。
- 20        これは、ヒトが持つ約300万の全SNPについてTaqMan法(注2)でタイピングをした場合のコストが約2億円(サンプル一人当たりの試薬コスト)で、SNP機能の統計的な解析に必要な数百から数千のサンプルについてこの全SNPのタイピングを行えば数百億円から数千億円の非現実的なコストと大規模な解析施設等のリソースが必要となるからである。
- 25        この為、通常のSNP機能解析プロセスでは、タイピングするSNP(以下、タイピングSNPと称す)を限定し、1千から1万程度のSNPに絞込んで機能

の解析を行う。

しかしながら、疾患易罹患性や薬剤応答性と SNP の関連の有無は、その SNP をタイピングした結果から統計的に解明する以外に方法はない。この為、最終的に関連が解明される「目標」 SNP (注 3) は、予め、タイピング SNP として 1 千から 1 万程度の SNP のグループに含まれて選定されていなければならない。これらの SNP が選定からもれた場合には、解析で関連のある SNP は見つからず、解析プロセスを再度タイピング SNP グループの選定からやり直さなければならない。

タイピング SNP を選び出す従来のやり方は、研究者が論文等の文献やゲノム関連のデータベース等を検索し、機能が既に解明しているヒト以外のゲノムと類似性したヒトの遺伝子の機能を予測するホモロジー検索等の手法を用いている。

しかしながら、これらのゲノム情報には、ヒト・ゲノムの機能が完全に記載されていない。この為、この SNP 機能解析プロセスの効率を決定するタイピング SNP を選び出すステップ、つまり如何に高い確率で「目標」 SNP を予測できるか否かは、研究者個人の経験とスキル、そして偶然の要素に大きく依存している。

また、SNP 機能解析プロセスの問題のもう一つにデータの品質がある。SNP 機能解析では、或る形質の発現の有無 (例、作用や罹患の有無) によって区分されるサンプル・グループの間で SNP のタイピングを行い、両グループにおける各 SNP のアレルの頻度を統計的に解析し、その形質を発現させる要因となる SNP を特定する。つまり、wet プロセスのタイピング・データに品質に問題があった場合、そのデータを基にした SNP 機能解析の結果は不正確なものになる。

この問題は、SNP のタイピングは人間が介在するプロセスであることに起因する。SNP 機能解析プロセス固有の品質の問題であるコンタミネーション、サンプルや試薬の取り違い等のオペレーション上のケアレス・ミス、これらデータ

## 3

品質低下の原因となる要素の多くが人的なものであり、これもまた研究者個人の経験とスキルに大きく依存している。

(注1) wetプロセスとは、SNPのタイピングを行うプロセス。現在のTaqMan法では、プレート上で血液等の遺伝子サンプルと試薬を反応・ハイブリタイゼーションさせ、その結果を光学的に測定し、最終的にサンプル個々がそのSNPで取るアレルのタイピングを行う。この工程をここではwetプロセスと称する。特定されたタイピングのデータの統計解析は、wetプロセスには含まれない。

(注2) 蛍光標識したアレル特異オリゴとTaq DNAポリメラーゼによるPCR (Polymerase Chain Reaction) 反応とを利用したタイピング方法。

(注3) 「目標」SNPまたは「目標」となるSNPとは、疾患易罹患性や(開発新薬の)薬剤応答性の要因となるSNP、及び疾患易罹患性や薬剤応答性の指標となるSNP、以上2つのいずれかに該当するものを意味する。SNP機能解析の目的は、これらのSNPを特定することである。

しかしながら、従来技術には以下に掲げる問題点があった。

タイピングの前に「目標」SNPを予測して、これらを含んだ数百から数千程度のSNPのグループを適切且つ確実に選択することは困難であり、また、SNPタイピング・データの品質を低下させるwetプロセス中の人的エラーであるサンプルや試薬の取り違いやコンタミネーションの発生を防止することも極めて困難であるという問題点があった。

本発明は斯かる問題点を鑑みてなされたものであり、その目的とするところは、①マーカーとなるSNPの推定とその近傍SNPの詳細なタイピングを繰り返すことで、「目標」SNPが存在すると思われる塩基配列領域を段階的に絞込、最終的に「目標」SNPを効率良く特定し、②wetプロセスにおけるデータ品質の要因となるオペレーション上のケアレス・ミスを防止するプロセス管理手法やコ

ンタミネーション等によって汚染されたデータを統計解析の前に排除することで、患者と非患者の統計量を比較し、SNP領域を絞り込むSNP特定方法に関する技術を提供する点にある。

## 5 発明の開示

請求の範囲第1項記載の本発明の要旨は、疾患易罹患性や薬剤応答性に関するSNP特定方法であって、SNPの解析の対象となる塩基配列領域において、予め走査領域を設定する第1のステップと、前記走査領域うち目標SNPを含む局所的な領域に段階的に絞込む第2のステップと、絞込まれた前記局所的な領域から前記目的SNPを特定する第3のステップとを有することを特徴とするSNP特定方法に存する。

請求の範囲第2項記載の本発明の要旨は、前記第2のステップは、前記目的SNPを特定するためのマーカーSNPを確定し、前記走査領域を段階的に絞込むステップを含むことを特徴とする請求の範囲第1項に記載のSNP特定方法に存する。

請求の範囲第3項記載の本発明の要旨は、前記第2のステップは、ハプロタイプ解析などの統計解析を用いて前記マーカーSNPを確定することを特徴とする請求の範囲第2項に記載のSNP特定方法に存する。

請求の範囲第4項記載の本発明の要旨は、前記第1のステップは、機能が解明されている遺伝子又は機能の予測ができる染色体等の限定されたゲノム領域において、前記塩基配列領域での走査領域を設定するステップを含み、前記第2のステップは、前記走査領域の中から、タイピングするSNPのグループ選定を行い、wetプロセスなどによるSNPタイピングを行う第4のステップと、前記SNPタイピングによるタイピング・データに基づき、前記走査領域において前記ハプロタイプ解析の各々組合せの出現する確率を統計量として求める第5のステップと、求められた前記統計量と予め設定又は測定された基準統計量とを比較し、

## 予め設

定された閾値を越える前記統計量と前記基準統計量との乖離がある場合、前記閾値を越えて乖離した位置に該当する領域に前記マーカースNPが含まれると判断する第6のステップとを含むことを特徴とする請求の範囲第3項に記載のSNP

5 特定方法に存する。

- 請求の範囲第5項記載の本発明の要旨は、前記第3のステップは、前記乖離が第1の閾値以下／未満の場合、前記第4のステップでのSNPのグループ選定におけるタイピングの対象となるSNP数を所定の割合で増加させ前記第5のステップを繰り返す第7のステップと、前記乖離が前記第1の閾値を越える／以上で、
- 10 且つ第2の閾値以下／未満の場合、乖離したピークの位置を含み、前記走査領域を所定の割合で縮小した新しい走査領域を設定して前記第5のステップを繰り返す第8のステップと、前記乖離が前記第2の閾値を越える／以上の場合、前記第2の閾値を越えて乖離した位置に該当する領域に前記マーカースNPが含まれると判断し、乖離したピークの位置を含み、前記走査領域を所定の割合で縮小した
- 15 新しい走査領域を設定して前記第5のステップを繰り返す第9のステップとを含むことを特徴とする請求の範囲第4項に記載のSNP特定方法に存する。

- 請求の範囲第6項記載の本発明の要旨は、前記第9のステップは、グループ選定されるSNPが所定数以下／未満になった場合、全DNAサンプルがタイピングされる目的SNPを含むSNPと確定するステップを含むことを特徴とする請求の範囲第5項に記載のSNP特定方法に存する。
- 20

請求の範囲第7項記載の本発明の要旨は、前記第7のステップは、前記第5のステップの処理が所定の回数を越えた／以上の場合、目標SNPが含まれないと判断して処理を中止するステップを含むことを特徴とする請求の範囲第5項に記載のSNP特定方法に存する。

- 25 請求の範囲第8項記載の本発明の要旨は、前記第8のステップは、前記第5のステップの処理が所定の回数を越えた／以上の場合、目標SNPが含まれないと

判断して処理を中止するステップを含むことを特徴とする請求の範囲第5項に記載のSNP特定方法に存する。

請求の範囲第9項記載の本発明の要旨は、前記第2のステップは、品質管理された工程でSNPのタイピングを行うステップを含み、前記品質管理された工程は、1つの試料に対して、1枚のアッセイ・プレートで4つのSNPタイピングを行い、カイ二乗検定などの統計手法による有意差があると認められたSNPタイピングの数が所定数を越える場合、タイピング・データは無効と判断して前記試料のコンタミネーションなど汚染されたデータを識別することを特徴とする請求の範囲第1項乃至請求の範囲第8項のいずれかに記載のSNP特定方法に存する。

請求の範囲第10項記載の本発明の要旨は、前記第2のステップは、前記有意差があると認められたSNPタイピングの数が所定数の場合、有意差が認められたSNPについてのみSNPタイピングを繰返し行い、有意差がなくなる結果が所定数続いたとき、前記タイピング／データを正しいと判定して採用することを特徴とする請求の範囲第9項に記載のSNP特定方法に存する。

請求の範囲第11項記載の本発明の要旨は、請求の範囲第1項乃至請求の範囲第10項のいずれかに記載のSNP特定方法における処理を実現可能なコンピュータプログラムであって、請求の範囲第1項乃至請求の範囲第10項のいずれかに記載の各ステップをコード化したことを特徴とするコンピュータプログラムに存する。

#### 図面の簡単な説明

第1図は、本発明の実施の形態1に係るSNP特定方法の概要を示すプロセス・フローである。

第2図は、第1図におけるステップ3、4及び6の詳細を示す図である。



第3図は、第1図のステップS5に関するサンプル・試薬チューブ、各種プレート及び関連したデータ・スキマの一例を示す図である。

第4図は、第3図のアッセイ・プレート10AP上におけるSNPの配置パターンの一例を示す図である。

5 第5図は、第1図のステップS7（段階5）の解析データの一例を示す図である。

第6図は、第1図のステップS7における解析データの一例を示す図である。

第7図は、従来のSNP機能解析のプロセス・フローを示す図である。

## 10 発明を実施するための最良の形態

以下、本発明の実施の形態を図面に基づいて詳細に説明する。

### （実施の形態1）

第1図は、本実施の形態1に係るSNP特定方法の概要を示すプロセス・フローである。第1図に示すように、本実施の形態1に係るSNP特定方法は、決定  
15 の対象となる開発新薬の決定（ステップS1）と解析するサンプルの収集（ステップS2）と「走査領域（塩基配列領域）」の決定（ステップS3）と「タイピング」SNPの決定（ステップS4）とwetプロセスによるSNPタイピング（ステップS5）とタイピング・データによるハプロタイプの解析（ステップS6）  
と「マーカー」SNPの推定（解析データの決定）（ステップS7）と「目標」SNP  
20 NP（目標SNP）の特定（ステップS8）とを有し、ステップS3～S7を1つのサイクル（処理サイクル）として繰り返す。

疾患易罹患性または薬剤応答性に関連するSNP特定を目的として、このプロセス・モデルでは、以下8つの工程（ステップ）を実施することによって、SNPのタイピングを行う「走査領域」から段階的に絞込んで、最終的に開発新薬の  
25 薬剤応答性の有無と関連する「目標」SNPを特定する。

8つのステップは以下のとおりである。まず、前段階1：決定の対象となる開

発新薬の決定を行う（ステップS 1）。次に、前段階2：解析するサンプルの収集を行う（ステップS 2）。段階1：「走査領域」の決定を行う（予め走査領域の設定する）（ステップS 3）。段階2：「タイピング」SNP（タイピングSNP）の決定を行う（ステップS 4）。段階3：wetプロセスによるSNPタイピングを行う（ステップS 5）。段階4：タイピング・データによるハプロタイプの解析を行う（ステップS 6）。段階5：「マーカー」SNPの推定を行う（解析データの決定）（ステップS 7）。段階6：「目標」SNPの特定を行う（ステップS 8）。この8つのステップにおいて、段階1から段階5（ステップS 3～S 7）までを1つのサイクル（処理サイクル）としてこれを繰り返す。

10 次に、第1図を参照して、各ステップにおける処理を詳しく説明する。

（イ）前段階1（決定の対象となる開発新薬の決定）：製薬企業等によって開発された新薬のうち、開発新薬の薬剤応答性（＝作用または副作用の有無）をSNPによって検定したものを選択する。また、このプロセス・モデルでは、例えば疾患易罹患性等を対象として、SNPとの関連を調べることもできる。

15 （ロ）前段階2（解析するサンプルの収集）：SNP機能解析では、或る形質の発現の有無によって区分されるサンプル・グループの間で、SNPデータを比較し、その形質を発現させる要因となるSNPを特定する（ステップS 2）。例えば、糖尿病等について疾患易罹患性とSNPとのを調べる場合には、糖尿病患者のグループと、Controlされたグループの間で、各SNPのアレル頻度を統計的に分類する。この時、Controlグループとして、糖尿病に罹患していないグループ、もしくは（糖尿病罹患の有無にかかわらず）無作為に抽出された平均的なサンプル・グループ、この2つのいずれかを用いる。

20 開発新薬の薬剤応答性に関連するSNPを特定する場合、作用もしくは副作用のあったグループと、これがなかったグループについて、段階1以降の解析を行う。

無作為に抽出された平均的なサンプル・グループが存在する場合、もしくはこ

れらに該当する外部のSNPデータが利用できる場合には、前述の2グループと合わせた3つのグループの間で、データの比較と解析を行うと、更に有効な解析を行うことが可能となる。

- (ハ) 段階1 (「走査領域 (塩基配列領域)」の決定) : このプロセス・モデルでは、この段階1から後に説明する段階5までを1つのサイクルとしてこれを繰り返すことによって、初期の大まかな「走査領域」からより局所的な「走査領域」へと段階的に絞込を行う。最後のサイクルでは、「走査領域」中の存在する全SNPのタイピング・データを解析することで、最終的に「目標」SNPを特定する。先ず、走査領域の決定を行う (ステップS3)。この「走査領域」とは、「目標」SNPの存在を調べる (走査する) 領域であり、ヒト・ゲノムの塩基配列上の連続した領域である。この領域の物理的な長さは、段階的に絞込まれるために可変である。

第2図は、第1図のステップS3、S4及びS6を説明する図である。第1図及び第2図を参照して、処理サイクルにおけるステップを更に、詳しく説明する。

- 第2図の(a)は、第1図のステップS3においてSNP10Dを含むゲノム領域の一例を示す。最初の「走査領域」は、ステップS3において、遺伝子やそれより大きな染色体等の大まかなレベルで規定 (設定) される。これは、現状でも染色体等レベルの大まかな機能が解明されているからである。また、複数の染色体が原因となり、どの染色体が怪しい (目的SNPを含むか) かわからない場合、ある特定の染色体を除いた残りの全ての染色体 (結果に男女の差が無い場合、性染色体は関係が無いので「走査領域」から除外するなどの措置をして) を対象とする場合、更に、極端な例として、全く目的SNPの絞込みに関する情報が無い場合など、全染色体を取りあえず「走査領域」として解析の対象とする方法も含む。また、これより詳細な、例えば遺伝子レベルで初期の「走査領域」を設定することもできる。即ち、予め機能が解明されている染色体レベルに基づき、走査領域 (1次走査領域、初期走査領域) を設定する。

2回目以降 ( $(n+1)$  回目、 $n$  は 1 以上の自然数とする) のサイクルにおいて、1 回目 ( $n$  回目) のサイクルにおける段階 5 (ステップ S 7) から再度この段階 1 (ステップ S 3) へ戻るので、 $n$  回目の処理サイクル ( $n$  次処理サイクル) における段階 5 で連鎖不平衡が確認された領域を、 $(n+1)$  回目の処理サイクル

5 (( $n+1$ ) 次処理サイクル) において新しい「走査領域 (( $n+1$ ) 次走査領域)」として設定する。この時、絞込まれた新しい「走査領域 (( $n+1$ ) 次走査領域)」は、前回 ( $n$  回目) のサイクルにおける「走査領域  $n$  次走査領域」の数分の 1 から数十分の 1 の長さとなる。なお、次のサイクルで、具体的にどの程度絞込むかについては、後に述べる SNP 間の連鎖不平衡の強度等によって異なる。

10 (二) 段階 2 (タイピング SNP の決定) : 段階 1 で設定された「走査領域」中から、「タイピング」する SNP のグループを選定する。

第 2 図の (b) は、第 2 図の (a) から選定された SNP のグループの一例を示す。これらグループに含まれる SNP は (遺伝子部位の機能やエクソン・イントロン等の区分に留意せず) 任意に選ぶことができるが、隣接する SNP 座位の間隔が出来る限り等間隔となるように選択を行う。これは、この一連の解析で SNP 座位間の連鎖不平衡を間接的に観測するため、連鎖不平衡に大きな影響を及ぼす SNP 座位間の物理的距離の差による誤差を排除するためである。

15 解析できる SNP 間の連鎖不平衡は、SNP 座位間の物理的距離が 1 ~ 10 万塩基程度と考えられる。この為、最初の「走査領域」が含まれる SNP が 10 万程度の染色体の全長である場合ならば、初回のタイピングは 1000 SNP 程度を行うことが望ましい。

2 回目以降の処理サイクルでは「走査領域」は段階的に絞込まれ、「走査領域」の物理的距離は短くなり、その範囲に含まれる SNP 数は減少する。タイピング SNP は、この「走査領域」の中から、数十から最大でも数百までの範囲で選択  
25 する。

第 3 図は、第 1 図のステップ S 5 に関するサンプル・試薬チューブ、各種プレ

ート及び関連したデータ・スキマの一例を示す図である。

(ホ) 段階3 (w e tプロセスによるSNPタイピング): 第1図のw e tプロセスによるSNPタイピングを行う(ステップS5)。即ち、選択されたSNPグループについて、各サンプルのSNPタイピングをT a qM a n PCR法等によって行う。このタイピング工程におけるコンタミネーションやサンプルの取り扱いによるデータの誤差を防ぐために、1) バーコードを用いたサンプル・試薬チューブ及びアッセイ用プレート等の世代管理、2) 「ハーディ、ワインバーグ平衡」を用いたタイピング・データの検定、これら1)、2)を行うことでタイピング・データの品質保証を行う。

- 10      1) バーコードを用いたサンプル・試薬チューブ及びアッセイ用プレート等の世代管理: タイピングの工程における最も基本的で多い間違いは、サンプルや試薬を間違えて取り扱うことである。現行のSNP解析のプロセスでは、最終的にアッセイを行うタイピング機器に使用するアッセイ・プレート10APを生成するまでにも、幾つかの中間プレートを生成する。その為、これらのプレートについて
- 15      いても正しいプレートから生成されているかが管理されることが重要である。

この為、サンプル及び試薬等のチューブ、これらから分注されて生成されるプレートについてバーコードを用いてID管理を行うと同時に、これらチューブ・プレート間又はプレート・プレート間の世代管理=どちらから(親)どちらが(子)生成されたかの関連を管理する。

- 20      サンプル: 下部に二次元バーコードを刻印されたサンプル・チューブ、及びこのサンプル・チューブが96本まで格納できるサンプル・ラックによって管理される。サンプル・ラック内の各サンプル・チューブの配置は、サンプル・ラックの底面からラック内の各サンプル・チューブのバーコードをスキャナーで読み取り、この配置をデータとして管理する。又、サンプル・ラック自身にもバーコード
- 25      ドが添付される。

試薬: 下部に二次元バーコードを刻印された試薬チューブ、及びこのサンプル・

チューブが96本まで格納できる試薬ラックによって管理される。サンプル・チューブと同様に、試薬ラック内の各試薬チューブの配置はバーコードをスキャナーで読み取り、この配置をデータとして管理され、試薬ラック自身にもバーコードが添付される。

- 5 プレート：プレートには、マスター・プレート、試薬プレート及びアッセイ・プレート10APの3種類がある。マスター・プレートは1枚につき96のウェル（サンプルや試薬を分注するプレート上の窪み）を持ち、同じく最大96のサンプル・チューブが収納可能なサンプル・ラックからサンプルが分注される。この時、1つのサンプルは1個所のウェル（ラック内のサンプル・チューブのレイ
- 10 アウトと同じポジションのマスター・プレートのウェル）に分注される。この為、マスター・プレートはサンプル・ラックの「子」のデータとして管理される。試薬プレートも（マスター・プレートと同様に）96のウェルを持つが、1枚の試薬プレートには1種類の試薬のみを使うので、試薬プレートは試薬チューブの「子」のデータとして管理される。試薬チューブの試薬ラック内の配置情報を保
- 15 持する理由は、分注に使う自動機器が、ラックに格納した試薬チューブの使用を要求するからである。アッセイ・プレート10APは384のウェルを持ち、最大96のサンプルに対して4 SNPのタイピングを同時に行うことができる。この1枚のアッセイ・プレート10APは、データ上4つの仮想的なバーチャル・プレートとして管理される。1つのバーチャル・プレートは最大96のサンプル
- 20 に対して1つのSNPのタイピングを行うプレートで、これが仮想的に4枚合わさり1つのアッセイ・プレート10APを構成する。

- 2) 「ハーディ、ワインバーグ平衡」を用いたタイピング・データの検定：SNPでは、その座位の塩基が2つのバリエーション、例えばA（アデニン）またはG（グアニン）の様な2つのアレル（対立遺伝子）を持つ。全ての染色体は対になっているので、SNP座位も対になっている染色体の各々に一つずつの合計2
- 25 箇所に存在する。この為、TaqManアッセイ等で観測されるSNPのパター

ンは、1) 一方アレル (この場合、A) のホモ接合体であるA-Aと、2) これとは異なるアレル (G) のホモ接合体G-G、3) 対立するアレルが一つずつ組み合わせあったヘテロ接合体A-G、以上の3つパターンになる。一方の遺伝子上でそのSNPがAを持つ確率を $\alpha$ とすれば、A-Aホモ接合体の確率は $\alpha^2$ であり、

5 G-Gホモ接合体の確率は $(1-\alpha)^2$ 、A-Gヘテロ接合体の確率は $2\alpha(1-\alpha)$ となり、この関係を「ハーディ、ワインバーグ平衡」と呼ぶ。「ハーディ、ワインバーグ平衡」が成立する条件は、多くの世代の交配を経て「平衡」した状態にあるサンプル集団から無作為に、同時に統計的に「平均」に抽出されなければならない。

10 ここで、そのサンプルのデータが「ハーディ、ワインバーグ平衡」から大きく離れた値を取った場合には、①そのデータを生成したアッセイ自体がコンタミネーション等によって「汚染」されたデータである。②タイピングされたサンプル集団が統計的には「無作為」に抽出されていない。①、②のうちのいずれかに相当すると考えられる。

15 SNP機能解析の目的は、或る形質の発現 (疾病等) の有無によって区分されるサンプル・グループの間でSNPデータを比較して、その形質を発現させる要因となるSNPを特定することである。つまり、要因となるSNPをタイピングした場合、そのグループを特徴付ける形質と因果関係のあるSNPを「期待される」以上に持っているため、「ハーディ、ワインバーグ平衡」から離れたアレルの

20 分布を持つはずである。言葉を代えれば、この2)「ハーディ、ワインバーグ平衡」を用いたタイピング・データの検定において、該当するSNPを見つけることこそ、SNP機能解析及び本プロセス・モデルの目的である。

「目標」SNPを特定するために、「ハーディ、ワインバーグ平衡」から逸脱したアレル分布のSNPから、①の「汚染」されたデータを持つSNPを除外する。

25 このコンタミネーション等の「汚染」されたデータを識別する方法が、1枚のアッセイ・プレート10APで4つのSNPをタイピングする方法である。

第4図は、第3図のアッセイ・プレート10AP上におけるSNPの配置パターンの一例を示す図である。384-ウェル（384のウェル）のアッセイ・プレート10APでタイピングされるSNPのレイアウトは第4図のようになる。

- 5 コンタミネーションやPCR法における不調等のアッセイ自体の失敗が起きた場合、同じプレート上の他のSNPについても異常が発生すると考えられる。

確率的には、要因となるSNPを偶然、複数選り出すことは極めて稀と考えられるので、同一プレート上で複数のSNPデータが「ハーディ、ワインバーグ平衡」から逸脱したケースは、タイピングの失敗と判断しそのプレートで行った全データの破棄（また、アッセイのやり直し）を行う。

- 10 更に、タイピングしたSNP同士が連鎖不平衡する可能性を回避するため、4つのSNPを可能な限り離す＝「走査領域」を4つの区間10PT（4つのウェルのブロック）に分け、その各々の区間からタイピングするSNPを1つ選択する。

- 15 第2図の（c）は、第1図のステップS6における走査領域の一例を示す。ウィンドウ10wを処理中のサイクルにおける「走査領域」の先頭から終端まで移動させて、そのウィンドウ10w内に含まれているSNPデータを解析するイメージを表している。

- （へ）段階4（タイピング・データによるハプロタイプの解析、ステップS6）：  
この工程では、ハプロタイプと連鎖不平衡の2つのコンセプトを用いて、「目標」  
20 SNPの近傍のSNP（最終的に「目標」SNP自身）を特定する。

一つの配偶子（対となる染色体の一方）上にある対立遺伝子（SNPのアレル）の組合せであるハプロタイプは、TaqMan法等のSNPタイピング・アッセイで得られたデータを基に、確率的に予測される。

- このハプロタイプを確率的に予測することについて、AまたはGを取るSNP  
25 #1、TまたはGを取るSNP#2、TまたはCを取るSNP#3、以上の3つのSNPが取るハプロタイプのケースから考える。



## 15

例えば、サンプルXで、SNP # 1がA-Aのホモ接合体、SNP # 2がT-Gのヘテロ接合体、SNP # 3がT-Cのヘテロ接合体を取る場合には、以下の2つのケースで4通りのハプロタイプの存在が予測される。

ケース 1)					確率
	SNP # 1	SNP # 2	SNP # 3		
5	染色体 1	A	T	T	25%
	染色体 2	A	G	C	25%
ケース 2)					
	SNP # 1	SNP # 2	SNP # 3		
	染色体 1	A	T	C	25%
	染色体 2	A	G	T	25%

10      なお、この例の各ハプロタイプの確率は各々25%となる。

更にもう一つのサンプルYが、SNP # 1がA-Aのホモ接合体、SNP # 2がT-Gのヘテロ接合体、SNP # 3がT-Tのホモ接合体を取る場合には、各々50%の確率を持つ2通りのハプロタイプが予測される。

ケース 1)					確率
	SNP # 1	SNP # 2	SNP # 3		
15	染色体 1	A	T	T	50%
	染色体 2	A	G	T	50%

そして、この2つのサンプルから予測されるハプロタイプは以下の通りとなる。

	SNP # 1	SNP # 2	SNP # 3	確率
	A	T	T	$25\% \div 2 + 50\% \div 2 = 37.5\%$
20	A	T	C	$25\% \div 2 = 12.5\%$
	A	G	T	$25\% \div 2 + 50\% \div 2 = 37.5\%$
	A	G	C	$25\% \div 2 = 12.5\%$

実際の解析は、数個から数十程度の範囲で定められた所定のSNP数を含む連続した領域をウィンドウ10wと定義し、そのウィンドウ10w内のSNPのタイピング・データ（全サンプル）から、ハプロタイプの組合せとその各々の出現する確率を統計的に求める。このウィンドウ10wに含まれるSNPの数が多す

ぎるとハプロタイプ個々の確率が低くなり、連鎖もしくは連鎖不平衡の有無の確認が微妙になるので、10程度のSNPを含むウィンドウ10wの定義が有効である。

第5図は、第1図のステップS7（段階5）の解析データの一例を示す図である。

第5図では、解析データにおける検定の前半部分にあたるハプロタイプ解析から統計量に変化が現れた領域を識別する手順を示す。

（ト）段階5（「マーカ」SNPの推定、ステップS7）：この工程では、第5図に示すように、特定のハプロタイプの確率が突出している領域を見つけ出すことによって、「目標」SNPの近傍領域の推定を行う。即ち、①WINDOW（ウィンドウ）10wを移動させながらWINDOW10w内のハプロタイプを解析する。②WINDOW10wの位置による「ハプロタイプ」数等の統計量の変化をプロットする。③比較するサンプル・グループ間で、統計量に顕著な差が認められる部分を抽出する。

段階4で解析されたハプロタイプのデータから、解析されたSNPグループ間で連鎖不平衡が見られたか否かの判定が可能となる。

これらのSNPの間に連鎖不平衡が見られない場合、各SNPのアレルの出現頻度は「平均的」な値に落ち着き、加えて、各々のSNPが「独立」しているため、そこから統計的に求められるハプロタイプについても、特定のハプロタイプに集約されない、広く薄く分散したものになると考えられる。

これに対し、解析されたSNPグループ間で連鎖不平衡が見られた場合、そこにはサンプル・グループを統計的に特徴付けるSNPが含まれていることであり、それらのSNPでは特定のアレルの出現頻度が増大する。そして、これらSNPデータを統計的に解析した結果であるハプロタイプの確率分布も、（「目標」SNPをアッセイしなかった場合より）特定のハプロタイプに集中することが予測される。

この特定ハプロタイプへの集中を識別する方法として、個々のハプロタイプの出現頻度を比較するほかに、その解析データから予測されるハプロタイプの総数、これらハプロタイプの標準偏差、確率上位のハプロタイプ・グループのハプロタイプ全体に対する出現頻度の割合を「統計量」として観測し、作用・副作用の有無という形質の発現によって区分されるサンプル・グループ間でこれらと比較する。

しかしながら、上述した様に、「目標」SNPを選び出して直接タイピングすることはなかなか困難である。この問題の解が、「目標」SNPが近傍のSNPと連鎖不平衡することを応用し、「目標」SNPの近傍領域を推定することである。連鎖不平衡した近傍のSNPは、「目標」SNPを直接解析した場合と比べれば程度こそ弱い、やはり同じようにハプロタイプの確率分布が変化することが期待される。このような近傍のSNPは、「目標」SNPの「マーカー」SNPであると考えられる。即ち、対象となるサンプル（効果のあったグループ：Caseグループ）の統計量と基準となるサンプル（効果のなかったグループ：Controlグループ）の基準統計量とを比較し、この差異が予め設定された閾値を越えた場合、該当するタイピング領域に変化があったと判断して（マーカーSNPと推定して）、該当するタイピング領域に対する所定の割合（例えば、5分の1～10分の1）を新しい走査領域に設定して次の処理サイクルを行う。

第6図は第1図のステップS7における解析データの一例を示す図である。第6図では、段階5の解析データの検定の後半部分にあたる新しい「走査領域」を定める手順を示す。

直接「目標」SNPをタイピングすること無しに、「マーカー」SNPのタイピング・データから、特定ハプロタイプの突出を示す領域を見つけ出し、その領域を次の「走査領域」の決定を行う段階1に戻り、再度工程を繰り返す。

又、そのサイクルで、「走査領域」中の全てのSNPがタイピングされているのであれば、次の段階6へ進んで、その特定ハプロタイプに収束しているSNP座

位が「目標」SNPとして決定される。

この段階の目的は「マーカー」SNPの推定であり、「走査領域」を新しく推定された「マーカー」SNPの近傍に絞込むサイクルを繰り返すことで、「マーカー」SNPを「目標」SNPを接近させることができる言える。なお、最終のサイクルの以前に「目標」SNPが解析から導き出されていたとしても、その時点ではまだそのSNPを「目標」SNPとして確定できてはいないので、「マーカー」SNPとなる。

(チ) 段階6（「目標」SNPの特定、ステップS8）：この工程の目的は、全工程で選り出された「目標」SNPと開発新薬の薬剤応答性との相関であり、その度合いを定量的に導き出すことである。

この工程では、「目標」SNPとして確定されたSNPのアレル頻度について、サンプル・グループでの最終的な検定を行う。（カイ2乗検定や最尤法等による）更に、サンプル・グループ以外のSNPデータ、特に東大医科研が所有する日本人の標準的なSNPに関するデータとの比較を行うことも効果的である。又、一部のゲノム研究企業で、SNPに関連するデータベースを販売する動きがあるので、これらの利用も可能である。

カイ2乗検定や相関解析を用いて、「～パーセントの確率で期待される効果が得られる」、または「～パーセントの確率で重篤な副作用が発生する」等の評価検定を行う。

ここで、データの汚染を識別する方法を詳しく説明する。

1) タイピング結果によって決定されたプレート中のSNPアレルの分布（1つのプレートでは最大96（実際には～92）のに対して4種類のSNPのタイピングを行うが、その各々のSNPについてのアレルの分布を意味する）

ここで「アレルの分布」とは、AA/AB/BBまたはAA/Aa/a a等の3値の何れかで表現されるSNPのタイピング・データを、そのグループ（同一プレートで同一SNPについてタイピングを行ったサンプルのSNPタイピング・

データ)において3値の何れを取るサンプルの合計数を示す。(例えば、AA : 23、AB : 54、BB : 15)

2) ハーディ・ワインバーグ平衡に基づいた理想的なSNPアレル分布:「ハーディ・ワインバーグ平衡」とは、「AA」の頻度を $\alpha^2$ とする場合、「BB」の頻度は $(1-\alpha)^2$ 、「AB」の頻度は $2\alpha(1-\alpha)^2$ 、となる状態を言う。

(この時、『「AA」の頻度』 > 『「BB」の頻度』)

(「A」単独の確率(=頻度)が $\alpha$ なので、両方とも「A」となる「AA」の頻度は $\alpha^2$ になり、「B」の確率=「A」じゃない確率は、 $1-\alpha$ 、「BB」の確率は $(1-\alpha)^2$ 、「AB」は「A」・「B」と「B」・「A」の2つのケースの合計の確率で $2\alpha(1-\alpha)^2$ になる。)

「ハーディ・ワインバーグ平衡に基づいた理想的なSNPアレル分布」とは、「AA」の頻度を $\alpha^2$ として、これから残りの「BB」の頻度と「AB」の頻度を、各々 $(1-\alpha)^2$ 及び $2\alpha(1-\alpha)^2$ と計算したもの。即ち、1)の例では、「AA」の頻度は $23/96=25\%$ で、「A」単独の確率は50%になる。この為、「BB」の頻度も25%でアレル分布は $96 \times 25\%=23$ 、「AB」の頻度は50%で分布は $96 \times 50\%=46$ が、「ハーディ・ワインバーグ平衡に基づいた理想的なSNPアレル分布」となる。

3) 上記2つのアレル分布の比較: 1)と2)の2つのSNPアレルが、カイ二乗検定などで「統計的に有意差」が有るか否かを確認する。

先の例では、AA AB BBが、

1) 23 54 15

2) 23 46 23

となり、ここでは、カイ二乗検定の結果は12.41%となり、「統計的に有意差」と認められる値(5%以下の確率)より大きい為、有意差はないものと判定され

る。

このようにデータの汚染が識別された後、

4) 比較した結果に有意差が認められた場合、そのプレート上で同時にタイピングされたSNPの2つ以上で3)の結果に「統計的に有意差」が認められた場合、そのプレート全体のタイピングは失敗であると判定し、そのタイピング・データを廃棄する。

- 5 5) 同一プレート上におけるSNPの内1つのSNPについてのみ「統計的に有意差」が認められた場合、その有意差が認められたSNPについてのみタイピングは再度実施される（同じ解析結果が得られた場合には、そのタイピング・データは「正しい」ものとして格納される。再実行されたタイピング結果に「有意差」がなくなれば、3回目のタイピングを行う。以下この手順を繰り返し、連続して
- 10 同じ結果が得られるまでタイピングを続ける。）

本実施の形態1に係るSNP特定方法は上記の如く構成されているので、以下に掲げる効果を奏する。

- タイピングされたSNPの中から疾患易罹患性または薬剤応答性に関連するSNPを特定する際に、解析の対象となる塩基配列領域を大まかな領域からより局
- 15 所的な領域へと段階的に絞込を行うことで、加えて、品質管理された工程でSNPのタイピングを行うことによって、最終的にこれら関連するSNPを特定することができる。

(実施の形態2)

次に、本実施の形態2に係るSNP特定方法を説明する。

- 20 段階1（第1図におけるステップS3）の「走査領域」の決定では、初期の「走査領域」を、大まかな染色体レベルより詳細な遺伝子レベルで設定することも可能である。これによって、狭い初期「走査領域」からタイピングが開始できるので、より効率的な「走査領域」が可能となる。

- 段階2（第1図におけるステップS4）の「タイピング」SNPの決定では、（特
- 25 に、早い段階の「走査領域」において）観測するSNP座位の間隔が大きすぎた場合、連鎖不平衡はほとんど観測することができない。この連鎖不平衡が観測可

能なレベルの物理的距離より短い間隔で「タイピング」SNPを選ぶことが必要である。

- 5 初期の「走査領域」では、この間隔を、1000 SNPについて1つのSNPを選ぶ。即ち、10万塩基について1つのSNPを選ぶように初期値が設定される。しかし、これはあくまで計算値であり、その近傍に見られる連鎖不平衡の度合いによっては、より小さい値（数十SNPについて1つのSNPを選ぶ）を取るケースもある。

- 10 段階5（第1図におけるステップS7）の「マーカ」SNPの推定において、ハプロタイプの総数や出現頻度等の「統計量」の変化を比較する以外の方法として、確率上位のハプロタイプが共通SNPパターンを求めることで連鎖もしくは連鎖不平衡の有無を確定することができる。

実施の形態1において説明したハプロタイプの確率の例では、サンプルXとサンプルYのハプロタイプの確率の和は以下の通りであった。

	SNP # 1	SNP # 2	SNP # 3	確率
15	A	T	T	$25\% \div 2 + 50\% \div 2 = 37.5\%$
	A	G	T	$25\% \div 2 + 50\% \div 2 = 37.5\%$
	A	T	C	$25\% \div 2 = 12.5\%$
	A	G	C	$25\% \div 2 = 12.5\%$

- 20 この例では、このハプロタイプのグループは、75%の確率でSNP # 1がAのアレルSNP # 3がTのアレルを取ることを示し、SNP # 1とSNP # 3に連鎖があったものと推測される。

	SNP # 1	SNP # 2	SNP # 3	SNP # 4	SNP # 5	確率
	A	T	T	A	T	30%
25	A	T	T	A	G	25%
	C	T	T	A	T	10%

22

C	T	T	A	G	10%
A	T	T	C	T	5%
A	T	T	C	G	5%
A	G	C	A	T	3%

- 5 55%の確率で、SNP # 1がアレルA、SNP # 2がアレルT、SNP # 3がアレルT、SNP # 4がアレルAを持つハプロタイプのパターンが観測される。更に、75%の確率でSNP # 2がアレルT、SNP # 3がアレルT、SNP # 4がアレルAを持つハプロタイプのパターンが、85%の確率でSNP # 2がアレルT、SNP # 3がアレルTを持つハプロタイプのパターンが観測される。
- 10 これによって、このサンプル・グループのデータからはSNP # 2とSNP # 3を中心とした連鎖が見られるものと判断される。

- ここで、連鎖の有無を分ける出現確率の閾値は、ウィンドウ10wで走査するSNPの数に関連する。SNP数が増えるとハプロタイプのバリエーションが増加して個々のハプロタイプが観測される確率が減るので、その結果、閾値の値も
- 15 低くなる。10のSNPを観測するウィンドウ10wを使った場合、閾値としては70%が適当と考えられる。なお、ウィンドウ10wで走査される領域（ウィンドウサイズ）は対象により変動するが、3-25 SNP程度が一般的に妥当と考えられる。

- 本実施の形態2に係るSNP特定方法は上記の如く構成されているので、実施
- 20 の形態1の奏する効果の他に以下に掲げる効果を奏する。

比較する2つのサンプル・グループ間に存在する異なった連鎖、もしくは一方のサンプル・グループにしか存在しない連鎖を確認することで、「目標」SNPまたはその近傍の領域を特定できる。

- なお、本実施の形態においては、ハプロタイプ解析を用いた説明を行ったが、
- 25 一般に知られている統計解析が利用できることは明らかである。また、本発明はそれに限定されず、本発明を適用する上で好適なSNP特定方法に適用すること



- ができる。また、本実施の形態を実現するため染色体レベル及びDNAレベルのSNPタイピング処理装置と、統計分析処理を行うコンピュータとを含み、一連処理ができるSNP特定システムも構成することができる。また、上記構成部材の数、位置、形状等は上記実施の形態に限定されず、本発明を実施する上で好適
- 5    5    な数、位置、形状等にすることができる。なお、各図において、同一構成要素には同一符号を付している。

#### 産業上の利用可能性

本発明は以上のように構成されているので、以下に掲げる効果を奏する。

- 10    以上説明したように、この発明によれば、マーカーとなるSNPを推定することで解析の対象となる塩基配列領域を大まかな領域からより局所的な領域へと段階的に絞込みを行い（患者と非患者の統計量を比較し、SNP領域を絞り込む）、更に、品質管理された工程でSNPのタイピングを行うことによって、最終的に、疾患易罹患性や薬剤応答性等に関連するSNPを特定することができる。

## 請求の範囲

1. 疾患易罹患性や薬剤応答性に関する SNP 特定方法であって、

SNP の解析の対象となる塩基配列領域において、予め走査領域を設定する第 1 のステップと、

5 前記走査領域うち目標 SNP を含む局所的な領域に段階的に絞込む第 2 のステップと、

絞込まれた前記局所的な領域から前記目的 SNP を特定する第 3 のステップとを有する

ことを特徴とする SNP 特定方法。

10 2. 前記第 2 のステップは、前記目的 SNP を特定するためのマーカー SNP を確定し、前記走査領域を段階的に絞込むステップを含むことを特徴とする請求の範囲第 1 項に記載の SNP 特定方法。

3. 前記第 2 のステップは、ハプロタイプ解析などの統計解析を用いて前記マーカー SNP を確定することを特徴とする請求の範囲第 2 項に記載の SNP 特定方法。  
15

4. 前記第 1 のステップは、機能が解明されている遺伝子又は機能の予測ができる染色体等の限定されたゲノム領域において、前記塩基配列領域での走査領域を設定するステップを含み、

前記第 2 のステップは、

20 前記走査領域の中から、タイピングする SNP のグループ選定を行い、wet プロセスなどによる SNP タイピングを行う第 4 のステップと、

前記 SNP タイピングによるタイピング・データに基づき、前記走査領域において前記ハプロタイプ解析の各々組合せの出現する確率を統計量として求める第 5 のステップと、

25 求められた前記統計量と予め設定又は測定された基準統計量とを比較し、予め設定された閾値を越える前記統計量と前記基準統計量との乖離がある場合、前記

閾値を越えて乖離した位置に該当する領域に前記マーカースNPが含まれると判断する第6のステップとを含む

ことを特徴とする請求の範囲第3項に記載のSNP特定方法。

5. 前記第3のステップは、

- 5 前記乖離が第1の閾値以下／未満の場合、前記第4のステップでのSNPのグループ選定におけるタイピングの対象となるSNP数を所定の割合で増加させ前記第5のステップを繰り返す第7のステップと、

前記乖離が前記第1の閾値を越える／以上で、且つ第2の閾値以下／未満の場合、乖離したピークの位置を含み、前記走査領域を所定の割合で縮小した新しい

- 10 走査領域を設定して前記第5のステップを繰り返す第8のステップと、

前記乖離が前記第2の閾値を越える／以上の場合、前記第2の閾値を越えて乖離した位置に該当する領域に前記マーカースNPが含まれると判断し、乖離したピークの位置を含み、前記走査領域を所定の割合で縮小した新しい走査領域を設定して前記第5のステップを繰り返す第9のステップとを含む

- 15 ことを特徴とする請求の範囲第4項に記載のSNP特定方法。

6. 前記第9のステップは、グループ選定されるSNPが所定数以下／未満になった場合、全DNAサンプルがタイピングされる目的SNPを含むSNPと確定するステップを含むことを特徴とする請求の範囲第5項に記載のSNP特定方法。

- 20 7. 前記第7のステップは、前記第5のステップの処理が所定の回数を越えた／  
以上の場合、目標SNPが含まれないと判断して処理を中止するステップを含む  
ことを特徴とする請求の範囲第5項に記載のSNP特定方法。

8. 前記第8のステップは、前記第5のステップの処理が所定の回数を越えた／  
以上の場合、目標SNPが含まれないと判断して処理を中止するステップを含む  
ことを特徴とする請求の範囲第5項に記載のSNP特定方法。

- 25 9. 前記第2のステップは、品質管理された工程でSNPのタイピングを行うステップを含み、前記品質管理された工程は、1つの試料に対して、1枚のアッセ

イ・プレートで4つのSNPタイピングを行い、カイ二乗検定などの統計手法による有意差があると認められたSNPタイピングの数が所定数を越える場合、タイピング・データは無効と判断して前記試料のコンタミネーションなど汚染されたデータを識別することを特徴とする請求の範囲第1項乃至請求の範囲第8項の

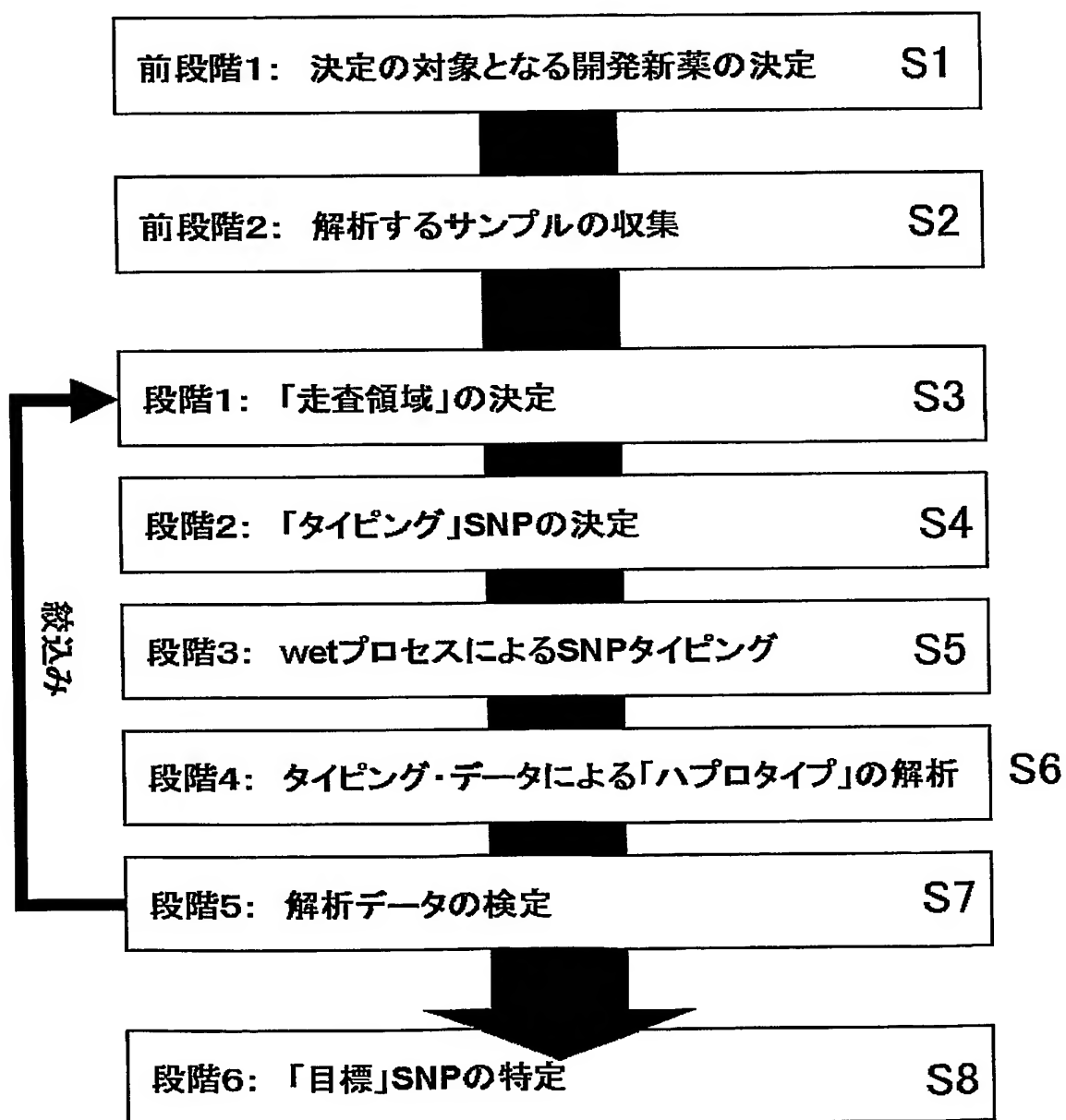
5 いずれかに記載のSNP特定方法。

10. 前記第2のステップは、前記有意差があると認められたSNPタイピングの数が所定数の場合、有意差が認められたSNPについてのみSNPタイピングを繰返し行い、有意差がなくなる結果が所定数続いたとき、前記タイピング／データを正しいと判定して採用することを特徴とする請求の範囲第9項に記載のS

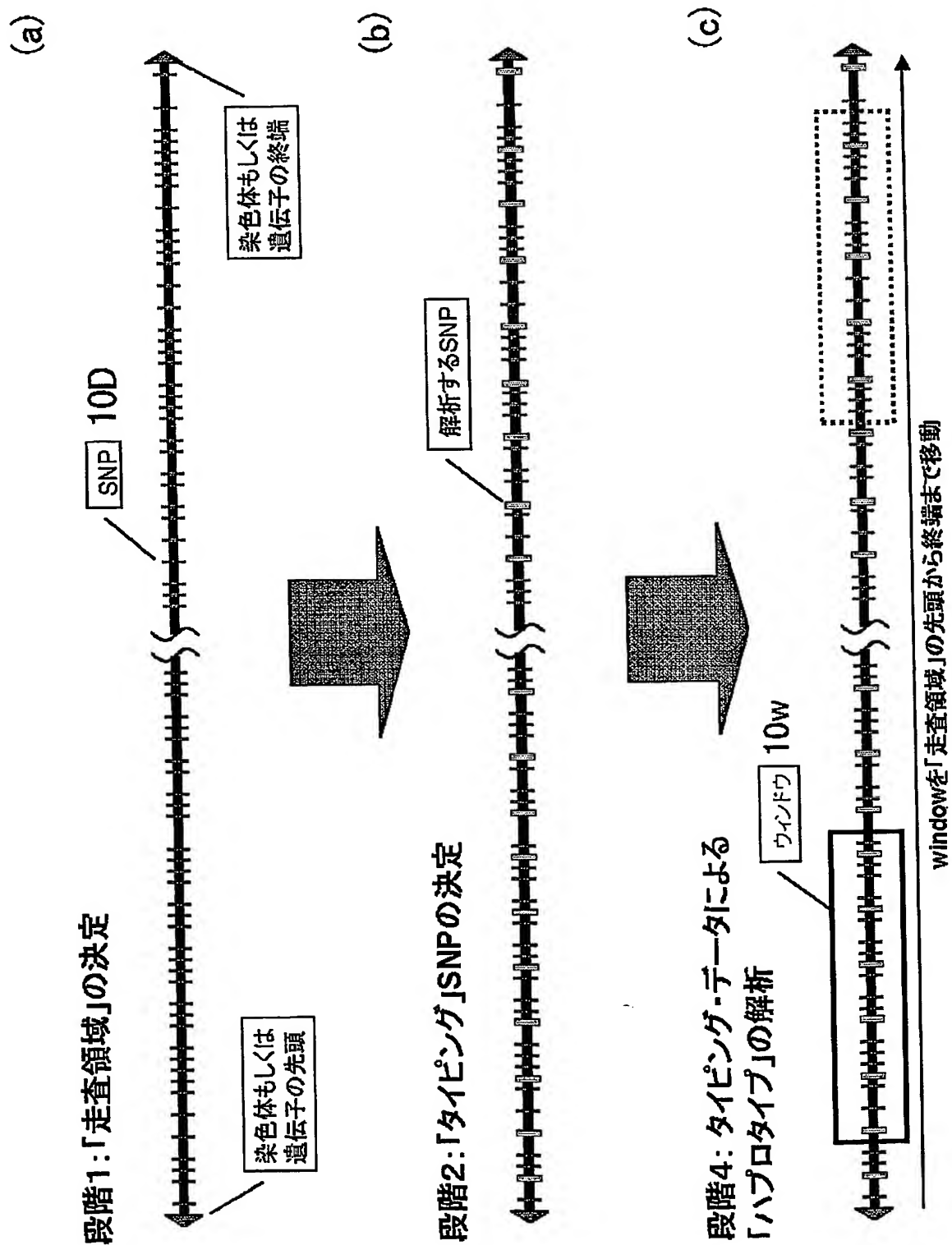
10 NP特定方法。

11. 請求の範囲第1項乃至請求の範囲第10項のいずれかに記載のSNP特定方法における処理を実現可能なコンピュータプログラムであって、請求の範囲第1項乃至請求の範囲第10項のいずれかに記載の各ステップをコード化したことを特徴とするコンピュータプログラム。

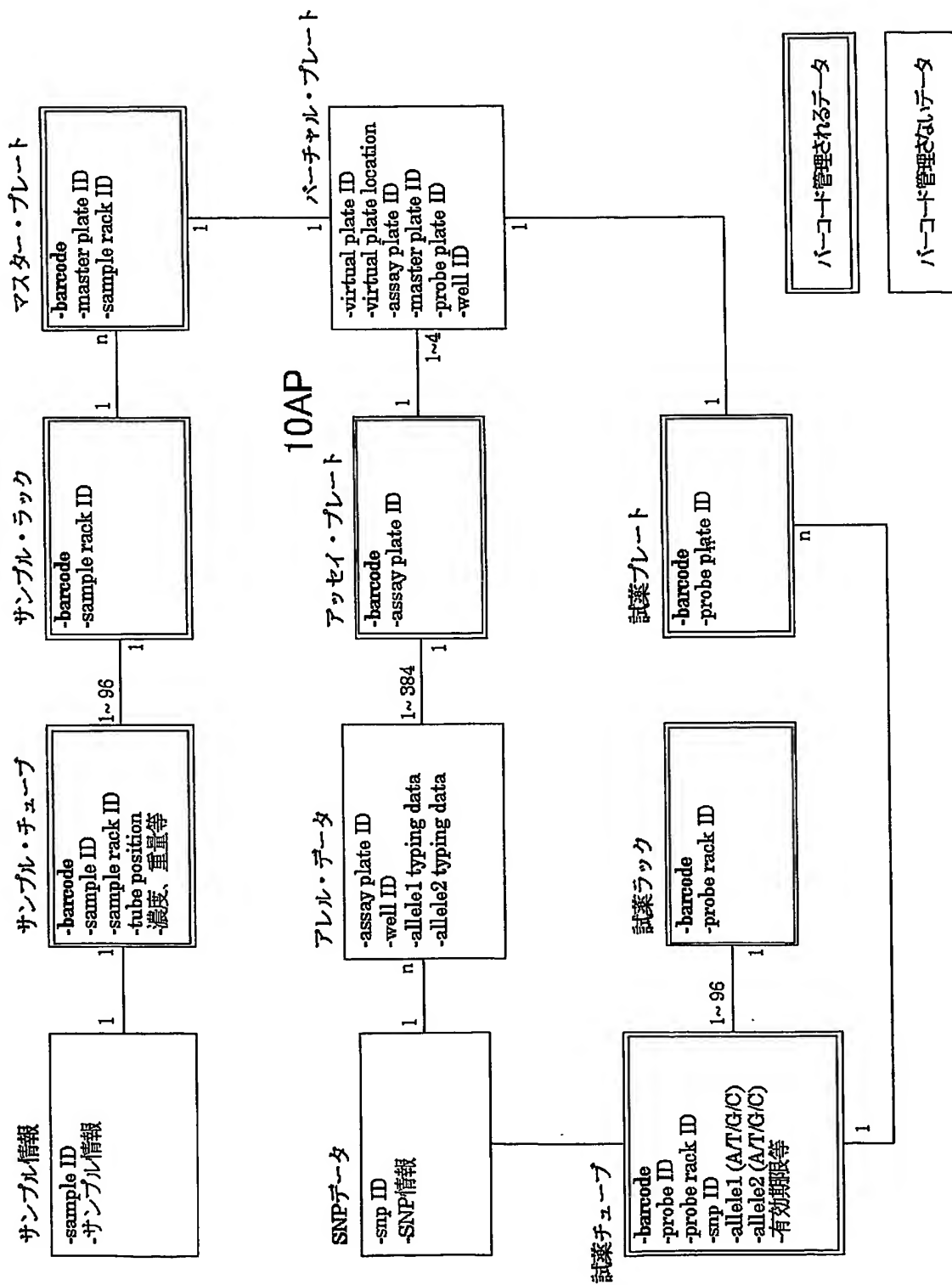
第1図

疾患易罹患性または薬剤応答性に関連するSNP特定のフロー

第2図



第3図



第4図

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
2	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
3	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
5	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
6	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
7	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
8	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
9	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
10	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
11	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
12	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
13	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
14	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
15	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
16	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
17	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
18	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
19	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
20	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
21	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
22	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4
23	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
24	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4

384ウェルのアッセイ・プレート上のSNPの配置

1	2
3	4

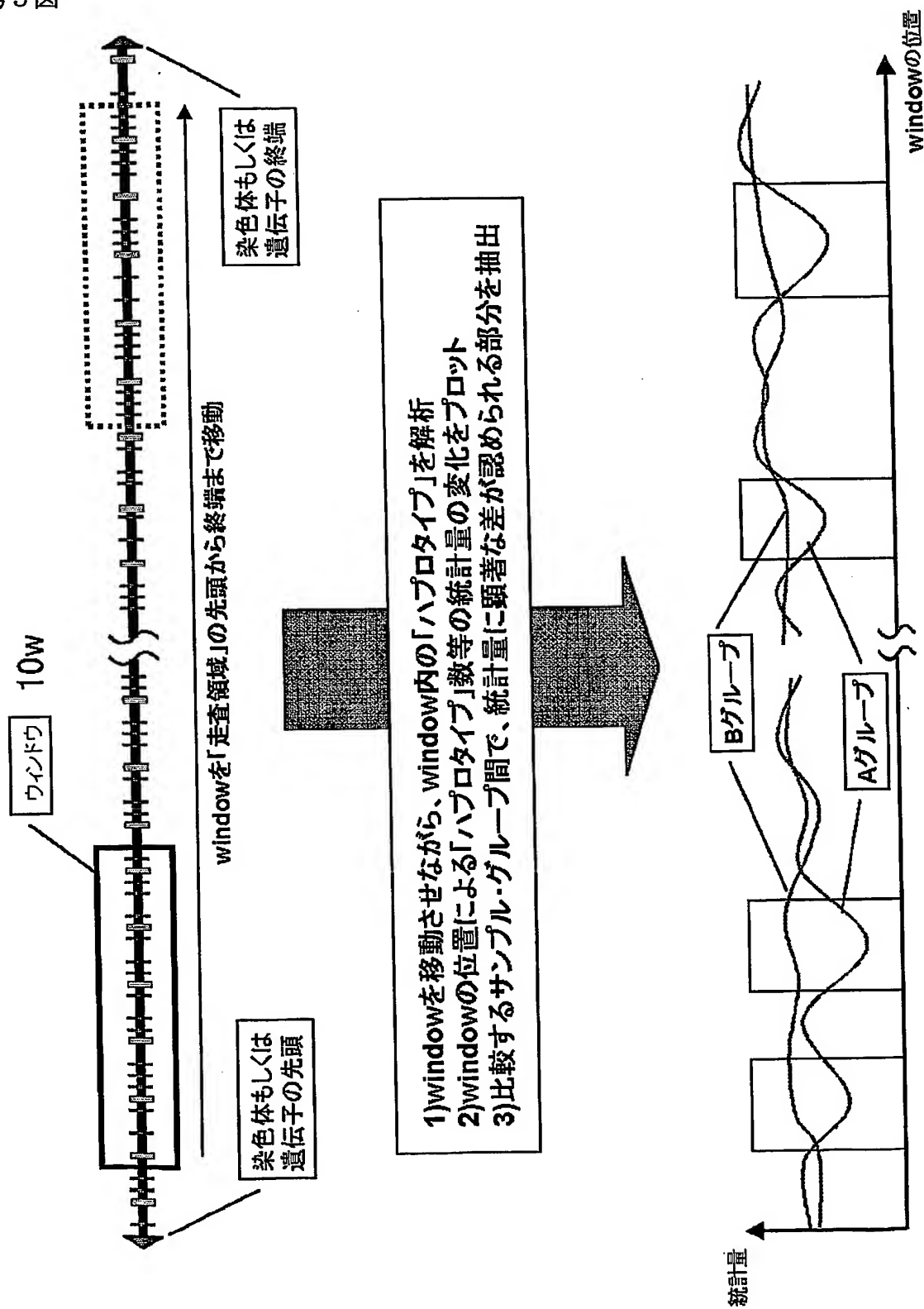
右図の 4つのウェルのブロックで、1つ

のサンプルに対し4種類のSNPをタイピング

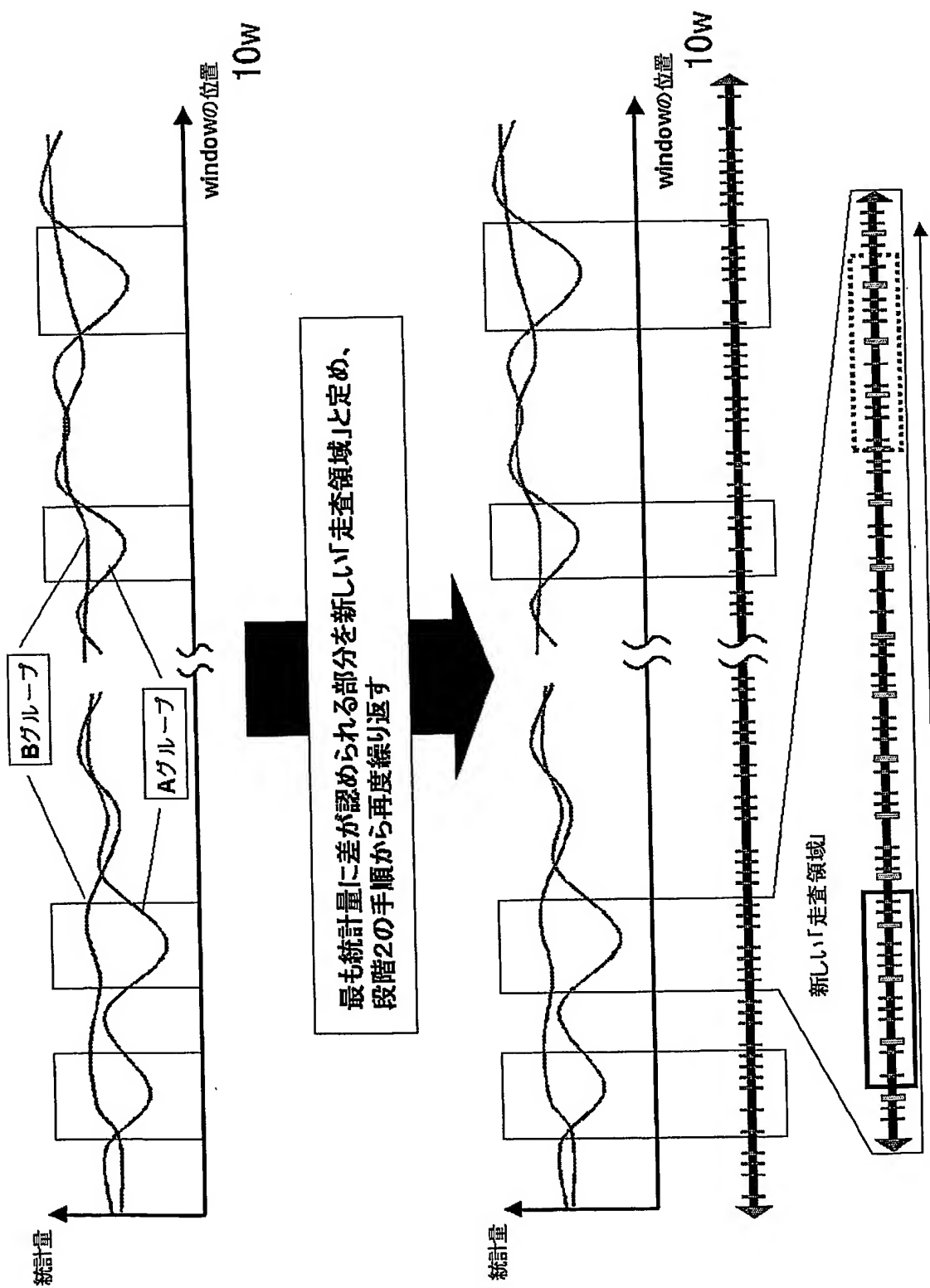
10PT



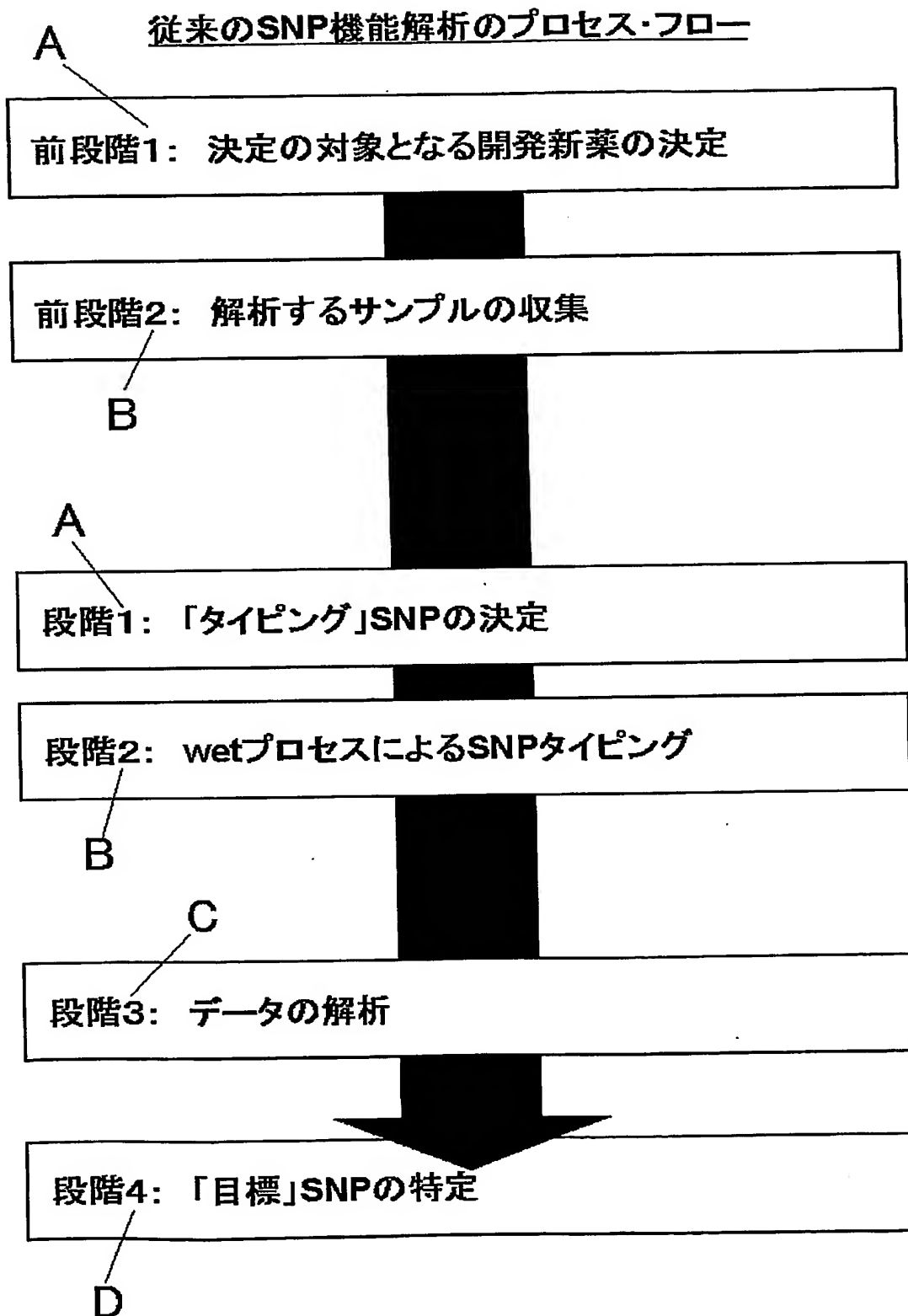
第5図



第 6 図



第7図



符号の説明

1 0 A P    アッセイ・プレート

1 0 D    SNP

1 0 P T    4つの区間

1 0 w    ウィンドウ (W I N D O W)

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/JP03/00261

## A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl<sup>7</sup> C12Q1/68, C12N15/09, G01N33/50

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl<sup>7</sup> C12Q1/68, C12N15/09, G01N33/50

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

BIOSIS/WPI (DIALOG), MEDLINE (STN), JICST FILE (JOIS)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Toshiaki INOUE, "Idenshi Tagata o Riyo shita Shikkan Idenshi Kensakuho, Post Sequence no Genome Kagaku (1) SNP Idenshi Tagata no Senryaku", 1st edition, Nakayama-Shoten Co., Ltd., 2000, pages 47 to 70	1-10
A	BARTON, A. et al., The single-nucleotide polymorphism lottery: how useful are a few common SNPs in identifying disease-associated alleles?, Genet Epidemiol. 2001, Vol.21, Suppl 1, p.S384-389	1-10
A	SAUNDERS, C.L. et al., Using single nucleotide polymorphisms to investigate association between a candidate gene and disease, Genet Epidemiol, 2001, Vol.21, Suppl 1, p.S415-420	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&amp;"

document member of the same patent family

Date of the actual completion of the international search  
12 February, 2003 (12.02.03)Date of mailing of the international search report  
25 February, 2003 (25.02.03)Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

**INTERNATIONAL SEARCH REPORT**

International Application No.

PCT/JP03/00261

**Box I Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 11

because they relate to subject matter not required to be searched by this Authority, namely:

This claim pertains to a mere "computer program" which is not specified by any concrete means in association with collaboration of software and hardware resources.

2. ☐ Claims Nos.:

because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:

because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**☐

The additional search fees were accompanied by the applicant's protest.

☐

No protest accompanied the payment of additional search fees.

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl. <sup>7</sup> C12Q1/68, C12N15/09, G01N33/50

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl. <sup>7</sup> C12Q1/68, C12N15/09, G01N33/50

最小限資料以外の資料で調査を行った分野に含まれるもの

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

BIOSIS/WPI (DIALOG), MEDLINE (STN), JICSTファイル (JOIS)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	井ノ上逸朗, 遺伝子多型を利用した疾患遺伝子検索法, ポストシーケンスのゲノム科学① SNP 遺伝子多型の戦略, 初版, 株式会社中山書店, 2000年, p. 47-70	1-10
A	BARTON, A. et al., The single-nucleotide polymorphism lottery : how useful are a few common SNPs in identifying disease-associated alleles?, Genet Epidemiol. 2001, Vol. 21, Suppl 1, p. S384-389	1-10

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの

「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」 口頭による開示、使用、展示等に言及する文献

「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&amp;」 同一パテントファミリー文献

国際調査を完了した日

12.02.03

国際調査報告の発送日

25.02.03

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

田村 明照

4 B

2936

電話番号 03-3581-1101 内線 3448

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	SAUNDERS, C.L. et al., Using single nucleotide polymorphisms to investigate association between a candidate gene and disease, Genet Epidemiol, 2001, Vol.21, Suppl 1, p.S415-420	1-10



第Ⅰ欄 請求の範囲の一部の調査ができないときの意見（第1ページの2の続き）

法第8条第3項（PCT17条(2)(a)）の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. ☒ 請求の範囲 11 は、この国際調査機関が調査することを要しない対象に係るものである。  
つまり、  
前記請求の範囲には、ソフトウェアとハードウェア資源が協働した具体的手段によつて特定されていない単なる「コンピュータ・プログラム」に係るものが記載されている。
2. ☐ 請求の範囲 \_\_\_\_\_ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. ☐ 請求の範囲 \_\_\_\_\_ は、従属請求の範囲であつてPCT規則6.4(a)の第2文及び第3文の規定に従つて記載されていない。

第Ⅱ欄 発明の単一性が欠如しているときの意見（第1ページの3の続き）

次に述べるようにこの国際出願に二以上の発明があるとの国際調査機関は認めた。

1. ☐ 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求の範囲について作成した。
2. ☐ 追加調査手数料を要求するまでもなく、すべての調査可能な請求の範囲について調査することができたので、追加調査手数料の納付を求めなかった。
3. ☐ 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったので、この国際調査報告は、手数料の納付のあった次の請求の範囲のみについて作成した。
4. ☐ 出願人が必要な追加調査手数料を期間内に納付しなかったので、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求の範囲について作成した。

追加調査手数料の異議の申立てに関する注意

- ☐ 追加調査手数料の納付と共に出願人から異議申立てがあった。
- ☐ 追加調査手数料の納付と共に出願人から異議申立てがなかった。

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☒ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**